NASA Technical Paper 1665

Rate Statistics for Radio Noise From Lightning

D. M. Le Vine, R. Meneghini,
and S. A. Tretter

MAY 1980

NASA

NASA Technical Paper 1665

Rate Statistics for Radio Noise From Lightning

D. M. Le Vine and R. Meneghini
Goddard Space Flight Center
Greenbelt, Maryland

S. A. Tretter
University of Maryland
College Park, Maryland

NASA

National Aeronautics
and Space Administration

Scientific and Technical
Information Office

1980

ABSTRACT

Radio frequency noise from lightning was measured at several frequencies in the HF – VHF range during the Thunderstorm Research International Project (TRIP) at the Kennedy Space Center, Florida. The data were examined to determine flashing rate statistics during periods of strong activity from nearby storms. It has been found that the time between flashes is modeled reasonably well by a random variable with a lognormal distribution.

Initially, the hypothesis that the occurrence of lightning flashes is a Poisson point process was tested using a uniform conditional test with Durbin's transformation and the Kolmogorov-Smirnov statistic. However, the Poisson hypothesis failed for the measured data. This resulted mainly from a lack of small time intervals, which we believe is a consequence of the finite duration of lightning flashes. This hypothesis has been tested by simulation, using a compound process in which the intervals between lightning flashes is assumed to be exponentially distributed and each flash is assumed to have a duration given by another independent random variable. Assuming flash rates and durations consistent with data, and that the flash duration has a fixed minimum of the order of the measuring system's resolution, the lognormal distribution consistently fit the simulated data.

CONTENTS

ILLUSTRATIONS

TABLES

RATE STATISTICS FOR RADIO NOISE FROM LIGHTNING

D. M. Le Vine and R. Meneghini
Microwave Sensors Branch
Goddard Space Flight Center
and
S. A. Tretter
Department of Electrical Engineering
University of Maryland

INTRODUCTION

The Goddard Space Flight Center (GSFC) investigates radiation from lightning as part of its program in severe storms research. The long-term goal of this research is to develop radiation from lightning as a remote sensing tool with which to monitor electrical processes in the atmosphere and hopefully to monitor the parent severe storms themselves (Christensen, et al. 1979). This research also provides the background necessary to address the problem of lightning hazards control (White and Haas, 1975; Le Vine, 1978).

The fact that NASA is an organization oriented toward space applications sets some implicit guidelines on the frequency range relevant for this research. In particular, interest is focused on radiation which penetrates the ionosphere (i.e., frequencies greater than a few MHz) which sets a lower bound, and because the radiated energy falls off rapidly with frequency (Horner, 1964; Kimpara, 1965; Oh, 1969), a practical upper bound is several hundred MHz. The GSFC program has focused on radiation in the range between 3 - 300 MHz. Relatively little is known about radiation in this frequency range (Pierce, 1977); consequently, a short-term goal of this work has been to fill the gaps in our understanding of the characteristics of the radiation and of the mechanisms responsible for its production.

The lightning flashing rate is one of several parameters of the radiation under investigation. The flashing rate is important in the design of systems such as communications systems and power distribution networks which are affected by radiation from lightning. It is a parameter of importance in the development of technology to detect lightning. It is a parameter of general interest in atmospheric science and it has even been suggested that the flashing rate may be indicative of the storm severity (Taylor, 1972 and 1973; Le Vine, 1976). GSFC has conducted research to provide information on the characteristics of the flashing rate during periods of strong activity from nearby storms.

DATA

The data to be presented here were obtained at the Kennedy Space Center as part of the Thunderstorm Research International Project (TRIP) and during an independent experiment prior to TRIP near Atlanta, Georgia during the late summer of 1975. The RF system used to collect the data consisted of several parallel channels, each tuned to a different frequency in the range between 3 and 300 MHz. Each channel was comprised of a vertically polarized antenna, filters, and an AM detector. The detector outputs from each channel were all recorded simultaneously on analogue tape, and the effective bandwidth of the entire system was 300 kHz (Le Vine and Krider, 1977; Le Vine et al., 1976). In addition to the RF radiation, slow electric field changes were also recorded. The slow electric field change is a low frequency, broadband measurement (near dc to 10's of kHz) which is indicative of the flash type (Uman, 1969).

A representative example of data is shown in Figure 1. Radiation at four frequencies is shown together with the slow electric field change (bottom trace). The data shown are a strip chart record of signals monitored during July 1976 at the Kennedy Space Center, Florida. Each group of dark vertical lines corresponds to a lightning flash. Data such as this were recorded during experiments at TRIP during the summers of 1976 and 1977, and also for a number of storms during the summer of 1975 near Atlanta, Georgia (Le Vine, et al., 1976). The data set includes several records of the complete growth-decay cycle of nearby storm cells. The flashing rate generally reflects the build-up-and-decay cycle typical of such cells (Byers and Braham, 1949). This is illustrated in Figure 2 for a storm monitored in July, 1976 at KSC, Florida during TRIP. The flashing rate data will obviously have trends as a result of the growth-decay cycle. We have attempted to compensate for these trends in a minimal way by partitioning the data into subsets of shorter duration. These subsets are referred to below as "tapes" and correspond to approximately 12 minutes of data.

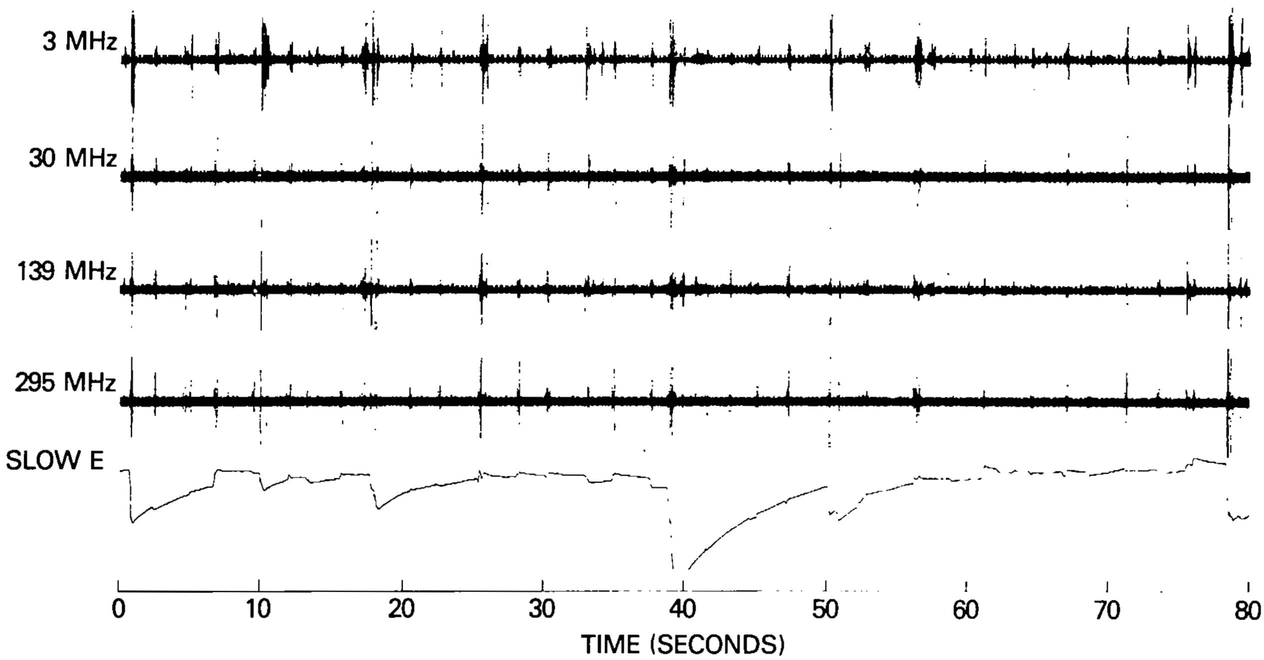


Figure 1. RF radiation from lightning. An example of RF radiation and slow electric field change data monitored on July 13, 1976 at the Kennedy Space Center, Florida during TRIP-76. The vertical scales are not calibrated.

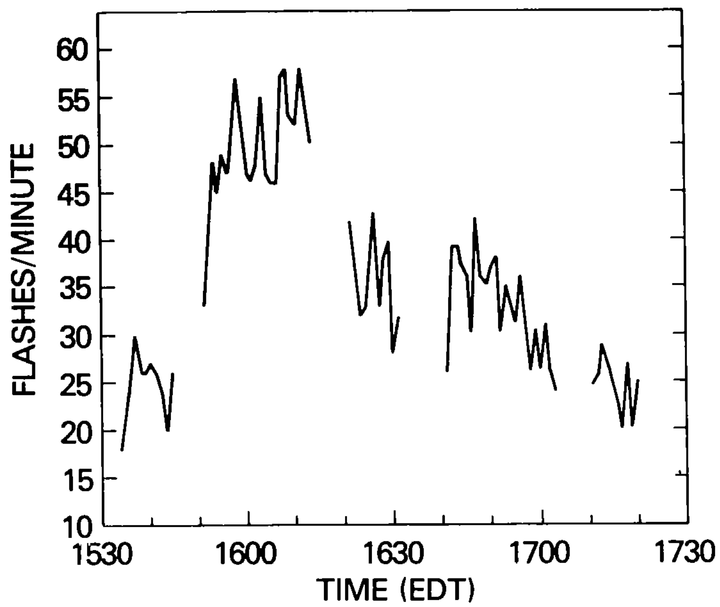


Figure 2. The lightning flashing rate for an isolated close cell. These data were obtained near Atlanta, Georgia on September 12, 1975.

The data have been analyzed to determine the probability density function and some moments for the time between flashes during periods of strong activity from nearby storms. The time interval between flashes was measured from records such as Figure 1 for four storms. A time interval was defined to be the time from the beginning of one flash until the beginning of the next. A flash was an event which appeared simultaneously at all four frequencies and exceeded an arbitrary but small threshold (i.e., peak amplitude). The hypothesis that the occurrence of lightning flashes was a Poisson point process was adopted initially. This hypothesis was tested, using a uniform conditional test with Durbin's transformation and the Kolmogorov-Smirnov statistic. However, the Poisson hypothesis failed for the measured data. Histograms of the time intervals suggest that the time between flashes is reasonably well modeled by a random variable with a lognormal distribution.

The Poisson process failed mainly from a lack of small intervals which we believe is a consequence of the finite duration of lightning flashes. Lightning is not a point process. In fact, each lightning flash is a complicated sequence of events lasting several tenths of a second. This is illustrated in Figure 3 which shows a representative cloud-to-ground flash with significantly more time resolution than in Figure 1. The event shown in Figure 3, a cloud-to-ground flash, would appear as a single cluster of dark vertical lines in Figure 1. In the analysis done here the time between lightning flashes was measured from the beginning of one flash to the beginning of the next. Such a measurement will be inherently lacking in small intervals because as the time between flashes gets to be on the order of the flash duration, the impulses in one flash will overlap those of the other; and because of the random nature of the impulses in a lightning flash, it is virtually impossible to distinguish two overlapping flashes from a single flash of more complex structure.

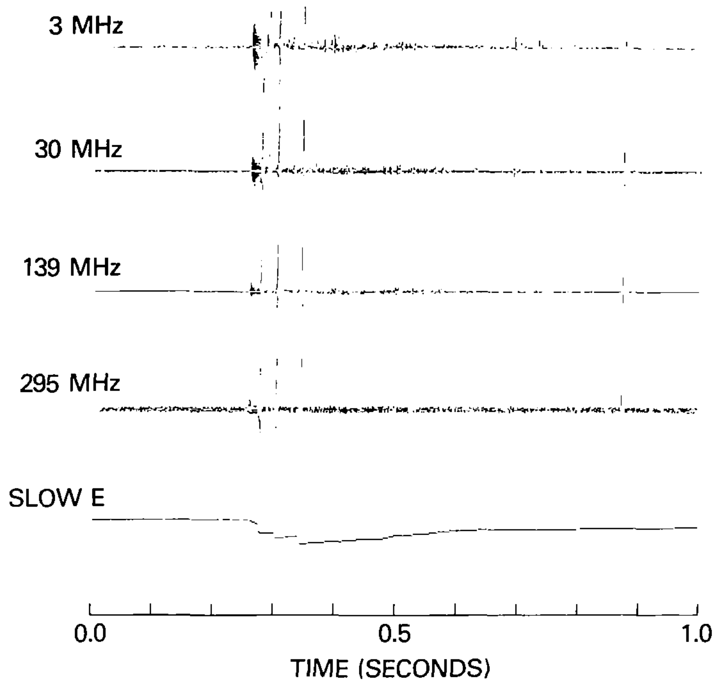


Figure 3. RF radiation and slow electric field changes of an individual cloud-to-ground lightning flash.

In light of this observation, it was conjectured that a Poisson process with overlaps may have produced the time between flash histograms that the lognormal density fit well. This hypothesis was tested by computer simulation, using a compound process in which the intervals between lightning flashes were allowed to be exponentially distributed, but each flash was assigned a duration given by another independent random variable. Assuming flash rates and durations consistent with data, and assuming that the flash duration has a fixed minimum representing the resolution of the measurement, the lognormal distribution consistently fit the simulated data. These results will be described in the following sections.

RESULTS

Analysis of the data began with the hypothesis that the time between events was Poisson. This seemed like a reasonable initial guess since the Poisson process has been successfully used to describe many physical point processes and because theoretically one would expect the time intervals to be exponentially distributed if the observed lightning were the result of pooling many independent point processes.

The Poisson hypothesis was tested using a computer program for the statistical analysis of series of events called SASE-V which is a modification of the program SASE-IV written by Lewis, Katcher and Weis (Lewis, et al., 1970). This is a Fortran program which implements the techniques described in the book *Statistical Analysis of Series of Events,* (Cox and Lewis, 1966). The length of the sequence of events that can be handled was reduced from 1999 to 1024 and the double precision arithmetic was eliminated to allow the program to run in a time-sharing mode on a UNIVAC 1108 computer. Otherwise, the capabilities of the two programs are identical.

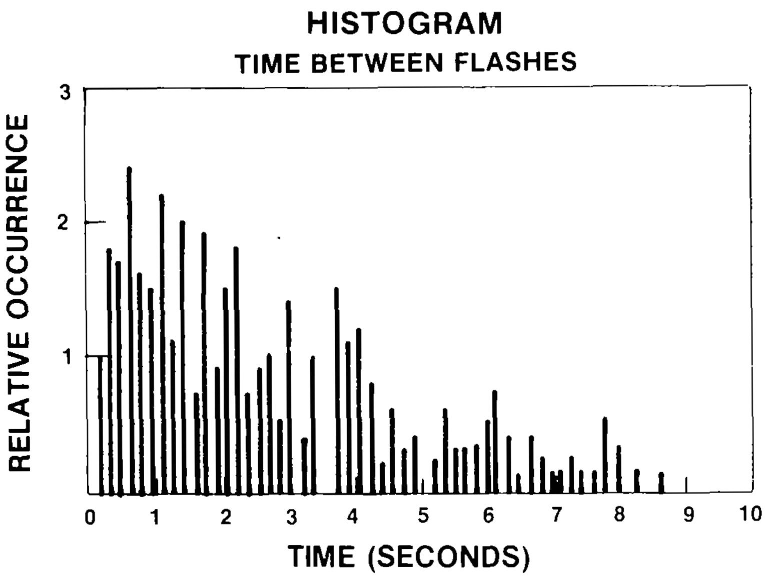
SASE-V performs several tests on the times between events. It computes some sample moments, the histogram, the log survivor function, the serial correlation function, the spectrum (periodogram) of the intervals, and the spectrum of the point process. In addition, it performs several uniform conditional tests, checking the time intervals and periodogram for goodness of fit to the Poisson hypothesis. These tests are described in more detail in Appendix A. Figure 4 shows an example of the results of several of these tests on data from a storm on July 13, 1976 at the Kennedy Space Center, Florida. In the case of a Poisson process, the histogram (upper left) would be exponential; the log survivor function (log of one minus the cumulative distribution) would be a straight line whose slope is the rate parameter of the Poisson process; the serial correlation coefficient would be approximately normally distributed about zero; and the moments (lower right) would have the characteristic values shown under the column "Poisson" in the table.

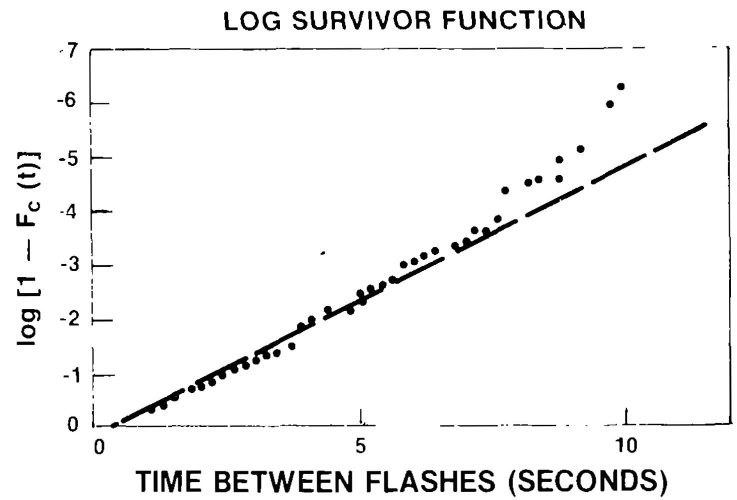
Many of the characteristics of the data were consistent, in a first approximation, with the Poisson hypothesis; however, the net result of SASE-V was to reject this hypothesis. In particular, in almost all cases, the Kolmogorov-Smirnov statistic for the uniform conditional test with Durbin's transformation strongly rejects the Poisson hypothesis.

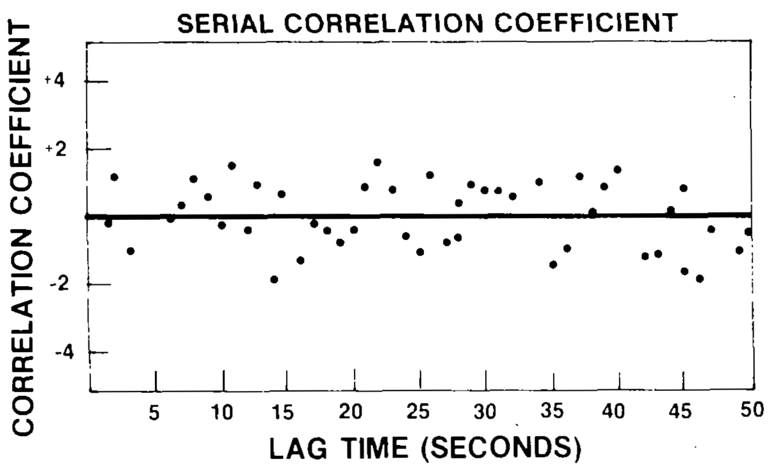
Some indication as to why the Poisson hypothesis failed can be obtained from a more detailed examination of the histograms. For the homogeneous Poisson process, the times between flashes should be exponentially distributed; however, the histograms show significant deviation from exponential, particularly near the origin. This is especially evident in Figure 5 which shows a histogram (from July 13, 1976) representative of the data. Notice that the density peaks near an interval of one second and then decreases rapidly to zero for smaller time intervals. The measurement accuracy (minimum time interval discernable on the strip charts) was on the order of 1/10 second and is represented by the data gap near the origin. The rapid decrease in intervals less than one second is a characteristic of the data and not a consequence of resolution. Deviations from exponential densities are also clearly apparent in the log survivor function (e.g., Figure 4) which shows deviations from linearity in the tails of the distribution.

On a more general level, the results indicate the relative merits of the uniform conditional test for Poisson processes with and without Durbin's transformation. They show that the test without Durbin's transformation has very weak power against many alternatives. The data clearly suggest that the uniform conditional test should not be used without Durbin's transformation (Tretter and Vaca, 1977). This is clearly evident in the column entitled "Exponential" in Table I, Appendix C, which shows the results of the uniform conditional test on the time intervals with (DN) and without (KS) Durbin's transformation.

The histograms clearly suggest distributions other than exponential, especially near the origin. Consequently, three alternative distributions (lognormal, Rayleigh and gamma) which are zero at the origin and have exponential-like tails were chosen for comparison with data. The three density functions are shown in Figure 6. For each partition of the data (i.e., "tape"), a maximum likelihood estimate of the parameters for each candidate probability density function was made and then







MOMENTS

	POISSON	DATA TAPE 1	DATA TAPE 2	DATA TAPE 3
MEAN	$1/\lambda$	3.5	2.9	1.8
STD DEVIATION	$1/\lambda$	3.1	2.4	1.6
COEF OF VARIATION	1	.9	.8	.9
SKEWNESS	2	2.1	1.4	2.0
KURTOSIS	6	8.9	5.0	7.9

8/5/75

Figure 4. Sample results of tests performed by SASE-V. Data was from July 13, 1976 at the Kennedy Space Center, Florida.

a Kolmogorov-Smirnov goodness-of-fit test for the distribution was made against the data (Appendix B). Table 1 of Appendix C shows the results of these tests for four storms (August 5 and September 12, 1975 in Atlanta, and July 13, 1976 at the Kennedy Space Flight Center, during TRIP). None of the choices is consistently accepted as a strong candidate. However, the Rayleigh distribution is strongly rejected in most cases. The gamma distribution function is occasionally strongly rejected. The lognormal density is rarely rejected and is accepted more often and at better levels than the others.

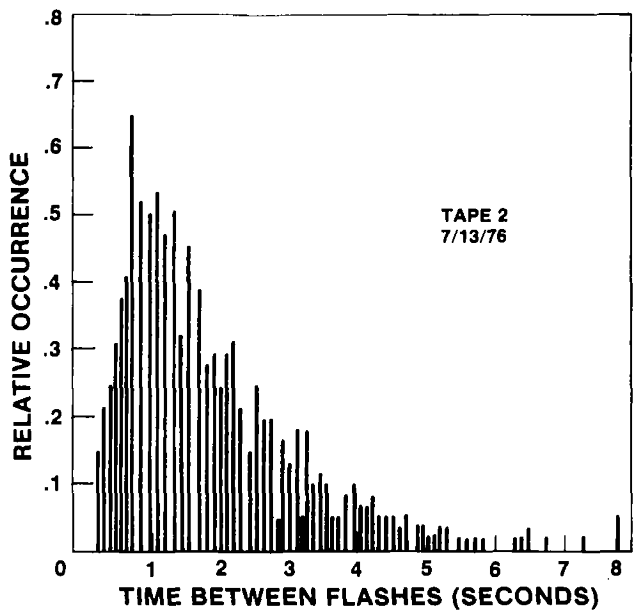


Figure 5. A representative example of the histogram for the time between flashes. These data are from tape #2, July 13, 1976

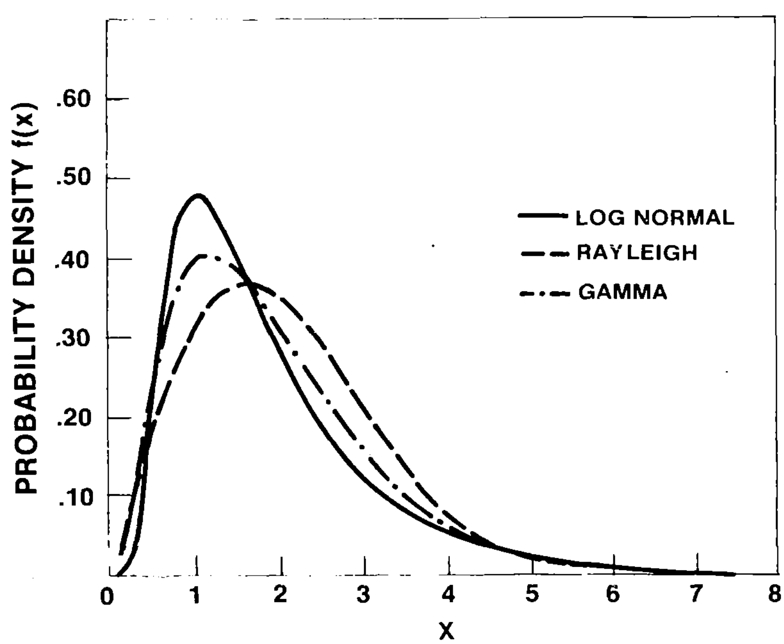


Figure 6. The probability density functions tested against the data.

Figure 7 is a plot of the best fit density functions against data from tape #2 of the July 13, 1976 storm and is representative of the results. The lognormal density appears to fit the data best because it is more peaked and because of its rapid decrease near the origin.

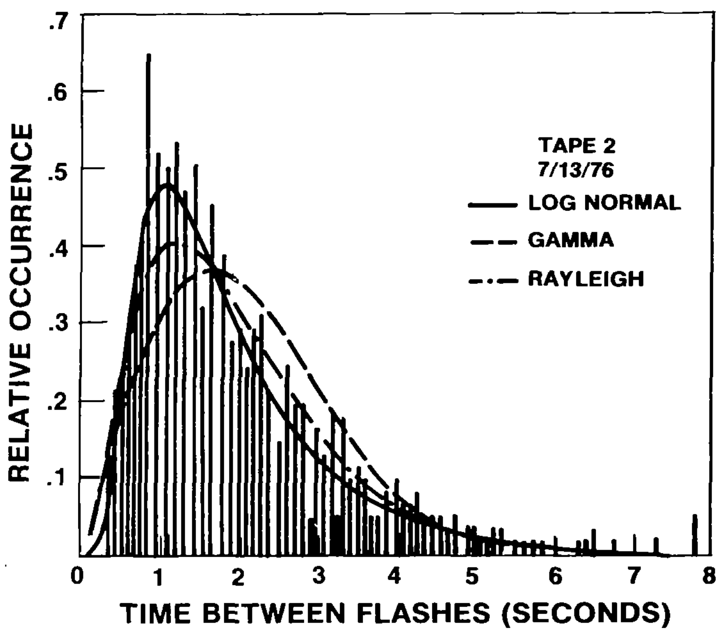


Figure 7. The best fit probability density functions and associated histogram. Data is from tape #2, July 13, 1976 (KSC, Florida).

ANALYSIS

On the basis of the results of SASE-V and examination of histograms, it seemed clear that a lack of small intervals contributes greatly to the failure of the Poisson hypothesis. The lack of small intervals is to be expected for events of finite duration such as lightning. This is so because when the time between events is small, overlaps can occur and with lightning it is not possible to distinguish overlapping flashes from individual flashes. In such a case, even if the occurrences of lightning flashes were in fact Poisson, the observed time intervals would not be exponentially distributed. It is possible that a lognormal density, as was observed for the measured data, is a close approximation to the true density function for such a process. That is, the results of SASE-V as described above may in fact not preclude a Poisson model for the occurrence of lightning. Determining the true density function for such a Poisson process with overlap is difficult analytically, and so it was decided instead to examine this problem using a computer simulation. The objective was to see if a Poisson process of finite events with overlaps resulted in histograms consistent with the observations of real lightning.

The principle behind the simulation is illustrated in Figure 8. The idea was to generate a marked process by associating with the occurrence times a new random variable for the duration of the flash. Thus, the computer generates sample values for two random variables, τ and σ. The τ_i represent the

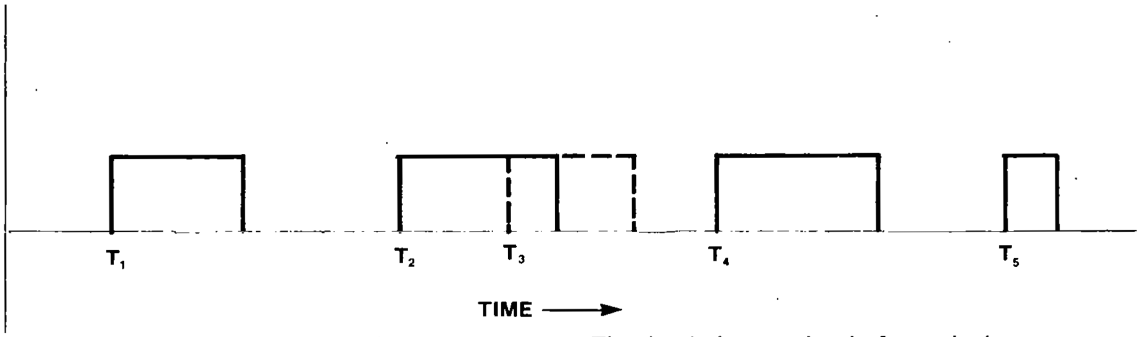


Figure 8. Simulated lightning occurrences. The simulation consisted of a marked process in which each occurrence is associated with another random variable for the duration. Overlapping events such as T_2 and T_3 become one event of longer duration.

time between the beginning of one lightning flash and the beginning of the next. To each τ_i is assigned a σ_i which represents the duration of the i-th flash. An example of a portion of the resulting sequence of events is shown in Figure 8. Having ordered the events as shown in the figure, a search is made for overlapping events (e.g., the second and third events in Figure 8). In the case of an overlap, the overlapping event becomes part of the original event (flash T_3 becomes an extension of T_2), and the time intervals, τ_i, are adjusted accordingly. The result is a set of $\hat{\tau}_i$ for a sample space with no overlaps and corresponding to the events which would be observed in experiments such as were actually performed in Atlanta and during TRIP.

The simulation was executed for several choices of τ and σ; however, the case in which the τ_i were exponentially distributed was of special interest. Figure 9 is the histogram resulting when the τ_i were exponentially distributed and the durations were modeled as a constant (representing the measurement resolution) plus an exponentially distributed random variable. Various combinations of rate parameters and constants were tested. Figure 9 is typical of the results obtained when rate parameters consistent with observed data were used. Clearly, Figure 9 is suggestive of the histograms obtained from measurements on real lightning (e.g., Figure 5). Other distributions such as uniform and Rayleigh were also tried for the τ_i. They resulted in density functions for the $\hat{\tau}_i$ which were not representative of the measured data.

To make the comparison more quantitative, histograms were produced from simulated data using a range of parameters representative of data (e.g., τ_i exponential, σ_i exponential plus a constant, and with rate parameters for each chosen to be consistent with data). The resultant simulation was then fitted with the gamma, Rayleigh and lognormal densities as had been done previously for the measured data (Appendix B). Tables 2–4, Appendix C, are summaries of the results.

In general, the Rayleigh density function was rejected consistently, as was true with the measured data; the gamma density generally was a best fit when the mean time between flashes was large and in fact it is the theoretical limit in the case of no overlaps; and the lognormal density fit best when the mean time between flashes was comparable to observed values (overlaps important). The results suggest that a Poisson process in which overlaps are important could result in time intervals distributed like those observed in our measurements on lightning.

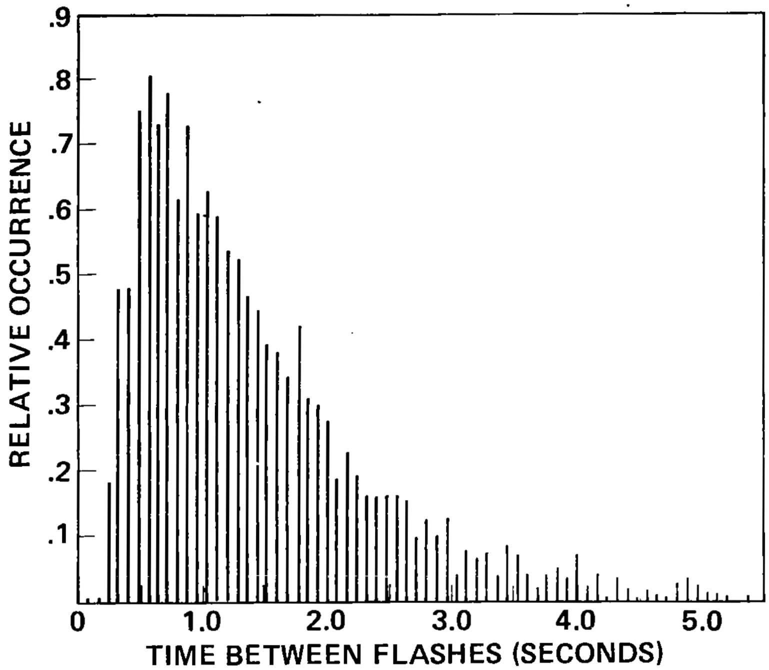


Figure 9. Histogram of the observed time between events obtained with the simulation. For the time between events, the simulation employed an exponentially distributed random variable and an exponentially distributed random variable plus a constant for the durations.

CONCLUSIONS

Our inferences are based on several storms, but more examples are really needed before unequivocal conclusions can be drawn. However, the available data and the comparison of simulated data and real data as compiled to date, suggest that the lognormal density is a reasonable choice with which to describe the time between lightning flashes for periods of strong activity. It also suggests that this density is consistent with the model of lightning as a compound Poisson process in which overlaps are important.

REFERENCES

1. Byers, N. R. and R. C. Braham, 1949, "The Thunderstorm," USGPO., Washington, D.C.

2. Christensen, L. S., W. Frost, W. W. Vaughan, 1979, "Proceedings: Workshop on the Need for Lightning Observations from Space," NASA CP-2095.

3. Cox, D. R. and P. A. W. Lewis, 1966, *The Statistical Analysis of Series of Events*, Methuen, London.

4. Horner, F., 1964, "Radio Noise from Thunderstorms," *Advances in Radio Research*, Vol 2, J. A. Saxton, Ed., Academic Press.

5. Kimpara, A., 1965, "Electromagnetic Energy Radiated from Lightning," *Problems in Atmospheric and Space Electricity*, pp. 352-365, S. C. Coroniti, Ed., Elsevier Publishing Company.

6. Le Vine, D. M., 1976, "Historical Review of the Association of Electrical Phenomena with Tornadoes," NASA TM-X-953-76-224.

7. Le Vine, D. M., H. H. Jenkins, B. J. Wilson and C. S. Wilson, 1976, "The Structure of Lightning Flashes HF-UHF," NASA TM-X-953-76-176.

8. Le Vine, D. M. and E. P. Krider, 1977, "The Temporal Structure of HF and VHF Radiation during Florida Lightning Return Strokes," Geophysical Research Letters, Vol 4 (#1), pp. 13-16.

9. Le Vine, D. M., 1978, "RF Radiation from Lightning," NASA TM-78095.

10. Lewis, P. A. W., A. M. Katcher and A. H. Weis, 1970, "SASE IV–An Improved Program for the Statistical Analysis of Series of Events," Program order number 360L-13.0.001, IBM Corp., Program Information Department, 40 Saw Mill River Rd., Hawthorne, New York, 10532.

11. Oh, L. L., 1969, "Measured and Calculated Spectral Amplitude Distributions of Lightning Sferics," IEEE Transactions on Electromagnetic Compatability, Vol EMC-11, (4), pp. 125-130.

12. Pierce, E. T., 1977, "Atmospherics and Radio Noise," *Lightning*, Vol 1, Chapter 10, R. H. Golde, Ed., Academic Press.

13. Taylor, W. L., 1972, "Atmospherics and Severe Storms," *Remote Sensing of the Troposphere*, Chap. 17, V. E. Derr, Ed., USGPO.

14. Taylor, W. L., 1973, "Electromagnetic Radiation from Severe Storms in Oklahoma during April 29-30, 1970," Journal Geophysical Research, Vol 78 (#36), pp. 8761-8777.

15. Tretter, S. A. and M. V. Vaca, 1977, "Experimental Performance of Uniform Conditional Goodness-of-fit Tests for Poisson Processes," Proceedings of Johns Hopkins Conference on Information Sciences and Systems, March 1977, p. 368.

16. Uman, M. A., 1969, *Lightning*, McGraw-Hill.

17. White, G. F. and J. E. Haas, 1975, *Assessment of Research on Natural Hazards*, Chapter 11, MIT Press.

APPENDIX A
SASE-V

Lewis, Katcher and Weis have written a FORTRAN program called SASE-IV for the statistical analysis of series of events (Lewis, et al., 1970). The program implements the techniques described in the book, *The Statistical Analysis of Series of Events*, by Cox and Lewis (1966). This program has been modified to run on the University of Maryland UNIVAC 1108 computer in the batch or demand modes and is called SASE-V. The principal modifications are the reduction of the length of the sequence of events that can be handled from 1999 to 1024 and the elimination of double precision arithmetic. Otherwise, the capabilities of SASE-V and SASE-IV are identical and are described in detail in Lewis, et al. (1970). The statistics computed and tests performed by SASE-V that have been found to be particularly useful in the analysis of lightning data are summarized in this section.

1. Sample Moments

Let $x_1, \ldots, x_N$ be a sequence of N positive numbers. Normally, these numbers would be the times between events in a point process. SASE-V computes the following sample moments and normalized moments:

(a) Sample mean

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

(b) Sample variance

$$\hat{\sigma}^2 \; \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \hat{\mu})^2$$

(c) Sample standard deviation

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

(d) Coefficient of variation

$$c = \hat{\sigma} / \hat{\mu}$$

(e) Third central moment

$$\hat{\mu}_3 = \frac{N}{(N-1)(N-2)} \sum_{n=1}^{N} (x_n - \hat{\mu})^3$$

(f) Coefficient of skewness

$$\hat{\gamma} = \hat{\mu}_3/\hat{\sigma}^3$$

(g) Fourth central moment

$$\hat{\mu}_4 = \frac{N(N^2 - 2N + 3)}{(N-1)(N-2)(N-3)} \sum_{n=1}^{N} (x_n - \hat{\mu})^4 - \frac{3(N-1)(2N-3)}{(N-1)(N-2)(N-3)} \hat{\sigma}^4$$

(h) Coefficient of Kurtosis

$$\hat{k} = \hat{\mu}_4/\hat{\sigma}^4 - 3$$

Many physically observed point processes are accurately modeled as homogeneous Poisson processes. A reasonable first step in the analysis of a point process is to see if the Poisson model applies. A simple check is to compare the sample moments defined above with the theoretical statistical moments. The times between events in a homogeneous Poisson process are independent, identically distributed random variables with an exponential probability density function, say $f(x) = \lambda e^{-\lambda x} u(x)$ where u (x) is the unit step function. In this case the ideal moments are:

(a) $\mu = E\{X\} = \lambda^{-1}$

(b) $\sigma^2 = E\{(X - \mu)^2\} = \lambda^{-2}$

(c) $\sigma = \lambda^{-1}$

(d) $c = \sigma/\mu = 1$

(e) $\mu_3 = E\{(X - \mu)^3\} = 2\lambda^{-3}$

(f) $\gamma = \mu_3/\sigma^3 = 2$

(g) $\mu_4 = E\{(X - \mu)^4\} = 9\lambda^{-4}$

(h) $k = \mu_4/\sigma^4 - 3 = 6$

2. The Sample Distribution Function, Log Survivor Function, and Histogram

SASE-V computes the sample distribution function for the sequence $\{x_n\}$ as

$$F_N(x) = \frac{\text{number of } x_n\text{'s} \leqslant x}{N}$$

This is an estimate of the actual distribution function $F(x) = P\{X \leqslant x\}$.

The function R (x) = 1 - F (x) is called the survivor function. SASE-V estimates R (x) as

$$R_N(x) = 1 - F_N(x)$$

It computes and plots the log survivor function

$$G_N(x) = \ln R_N(x)$$

For the Poisson process, G (x) = ln R (x) = $-\lambda x$. Thus deviations from linearity in the plot of $G_N(x)$ give insight into how the observed processes differ from an ideal Poisson process. Another closely related function is the hazard function

$$h(x) = -\frac{d}{dx} G(x) = \frac{f(x)}{1 - F(x)}$$

The hazard function can be estimated from the plot of the log survivor function. For the ideal Poisson process, h (x) = λ.

SASE-V makes a direct estimate of the probability density function for the random variables $\{X_n\}$ by computing and plotting a histogram for the observations $\{x_n\}$.

3. Goodness-of-Fit Tests for the Poisson Process

SASE-V performs several goodness-of-fit tests to check if the homogeneous Poisson process hypothesis is statistically reasonable. All of these tests are based on the fact that if a homogeneous Poisson process is observed over the interval (0, T), the observed normalized arrival times

$$y_i = t_i/T = \sum_{n=1}^{i} x_n/T \ , \ i = 1, \ldots, N$$

are the order statistics of a random sample of size N from a population uniformly distributed over (0, 1). Tests based on this fact are called uniform conditional tests. These tests are particularly convenient for the Poisson hypothesis because the rate parameter λ need not be estimated and no grouping of data is required.

Let the sample distribution function for the normalized arrival times be

$$F_N(y) = \frac{\text{number of } y_i \leqslant y}{N}$$

For the homogeneous Poisson process hypothesis, the theoretical distribution function would be

$$F(y) = \begin{cases} 0 \text{ for } y \leqslant 0 \\ y \text{ for } 0 \leqslant y \leqslant 1 \\ 1 \text{ for } y \geqslant 1 \end{cases}$$

SASE-V computes the following four statistics:

(a) The One-Sided Kolmogorov-Smirnov Statistics

$$D^+_N = \sqrt{N} \sup_{0 \leqslant y \leqslant 1} [F_N(y) - y] = \sqrt{N} \max_{1 \leqslant i \leqslant N} [\frac{i}{N} - y_i]$$

$$D^-_N = \sqrt{N} \sup_{0 \leqslant y \leqslant 1} [y - F_N(y)] = \sqrt{N} \max_{1 \leqslant i \leqslant N} [y_i - \frac{i-1}{N}]$$

(b) The Two-Sided Kolmogorov-Smirnov Statistic

$$D_N = \sqrt{N} \sup_{0 \leqslant y \leqslant 1} |F_N(y) - y| = \max \{D^+_N , D^-_N\}$$

(c) The Anderson-Darling Statistic

$$W^2_N = N \int_0^1 \frac{[F_N(y) - y]^2}{y(1-y)} dy$$

$$= -N - \frac{1}{N} \sum_{i=1}^{N} \{(2i-1) \ ln\, y_i + [2(N-i) + 1] \ ln\,(1-y_i)\}$$

Each of these statistics gives a measure of the deviation of $F_N(y)$ from the hypothesized distribution $F(y) = y$. The Anderson-Darling statistic emphasizes deviations near 0 and 1 as a result of the factor $y(1-y)$ in the denominator.

These statistics are used by selecting a threshold z and rejecting the hypothesis H_0, that the process is a homogeneous Poisson process, if the threshold is exceeded. The threshold is selected so that the test has a desired level α, that is, so that

$$P \{ \text{statistic} > z \mid H_0 \} = \alpha$$

Asymptotic formulas for the levels of the Kolmogorov-Smirnov statistics are derived in Kendall and Stuart (1961). The asymptotic formula for the two-sided Kolmogorov-Smirnov statistic is

$$\lim_{N \to \infty} P\left\{D_N > z\right\} = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp(-2r^2z^2)$$

The approximation is satisfactory for $N \geqslant 80$. Values of this expression for a range of z are given in Table 5, Appendix C. A table of asymptotic levels for the Anderson-Darling statistic is given by Lewis (1961).

The power of a test is defined to be the probability that the null hypothesis H_0 is rejected given that an alternative hypothesis H_1 is true. Tests for the null Poisson hypothesis based on the uniform conditional property are not very powerful against a variety of alternatives. For example, suppose that the normalized arrival times $\{y_i\}$ are equally spaced in the interval (0, 1). Then $F_N(y)$ remains close to $F(y) = y$. Durbin (1961) has suggested a modification of the uniform conditional tests that give a large increase in power over the uniform conditional tests for a broad class of alternatives. The first step in Durbin's modification is to order the sequence of times between events $\{x_n\}$ to obtain the observed order statistics

$$0 < x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(N)}$$

The next step is to compute the sequence

$$w_i = \frac{x_{(1)}}{t_N} + \frac{x_{(2)}}{t_N} + \ldots + \frac{x_{(i-1)}}{t_N} + (N+2-i)\frac{x_{(i)}}{t_N} \quad \text{for } i=1, \ldots, N-1$$

where

$$t_N = \sum_{n=1}^{N} x_n$$

It can be shown that under the null Poisson hypothesis, the sequence $\{w_i\}$ has the same distributional properties as the sequence $\{y_i\}$. Thus, the Kolmogorov-Smirnov and Anderson-Darling tests described above can be applied to $\{w_i\}$. The statistic, D_N, computed by SASE-V has been labelled KS when Durbin's transformation is not used and DN when Durbin's transformation is used.

4. Serial correlation coefficients

SASE-V computes the serial correlation coefficients

$$\hat{\rho}_j = \frac{N}{N-j} \frac{\sum_{i=1}^{N-j} (x_i - \hat{\mu})(x_{i+j} - \hat{\mu})}{\sum_{i=1}^{N} (x_i - \hat{\mu})^2}$$

for j = 1, 2, . . . , min (N/2, 100) where $\hat{\mu}$ is the sample mean. This sequence is an estimate of the autocovariance function

$$\rho_j = \text{cov}\,(X_i, X_{i+j}) \,/\, \text{var}\,(X_i)$$

Under the hypothesis that the x_n's are uncorrelated, that is, $\rho_j = 0$ for $j \neq 0$, the $\hat{\rho}_j$'s are approximately normally distributed with zero mean and variance $(N\text{-}j)^{-1/2}$. This approximation is reasonable for $N \geqslant 100$ if the skewness is moderate. SASE-V plots $(N\text{-}j)^{1/2}\,\hat{\rho}_j$ as a function of j.

A renewal process is a point process in which the times between events are independent, identically distributed random variables. The homogeneous Poisson process is a special type of renewal process. Independent random variables in a set are always uncorrelated. However, a set of uncorrelated random variables may or may not be independent. Thus, the serial correlation coefficients can indicate whether or not a renewal process model is appropriate.

5. Spectrum of the Intervals

SASE-V estimates the spectral density of the sequence of times between events as

$$\hat{S}(w) = \frac{1}{\pi}\left\{1 + 2\sum_{n=1}^{M} \hat{\rho}_n\, a_n \cos(nw)\right\}$$

where a_n is the Parzen window, that is,

$$a_n = \begin{cases} 1 - 6\left(\dfrac{n}{M}\right)^2 + 6\left(\dfrac{|n|}{M}\right)^3 & \text{for } |n| \leqslant M/2 \\[2ex] 2\left(1 - \dfrac{|n|}{M}\right)^3 & \text{for } \dfrac{M}{2} \leqslant |n|\, M \\[2ex] 0 & \text{elsewhere} \end{cases}$$

It also computes the discrete Fourier transform

$$B_k = \sum_{n=1}^{N} x_n\, e^{-j2\pi(n-1)k/N} \quad \text{for } k=0, \ldots, N\text{-}1$$

and the periodogram

$$C_k = |B_k|^2/N \quad \text{for } k=0,\ldots,N-1$$

It can be shown that if the times between events are independent identically distributed random variables, then the values of C_k for $0 < k < N/2$ are asymptotically independent, identically distributed random variables with the exponential density. Consequently, the goodness-of-fit tests described in Section 3 can be applied to the periodogram to check the renewal process hypothesis. SASE-V applies the Kolmogorov-Smirnov and Anderson-Darling uniform conditional tests to the periodogram.

6. Spectrum-of-the-Point Process

SASE-V estimates the spectrum-of-the-point process in the following way:

Let

$$D = (t_N - t_1)/(N-1)$$

$$A(k) = \sum_{i=2}^{N} \exp\,[j\,B\,(t_i - t_1)/D]$$

and

$$I(k) = 2\,|A(k)|^2\,/(N-1)$$

for k = 1, 2, . . . , P. B is normally chosen as $2\pi/(N-1)$. P should be chosen to be larger than N. Usually all the important features of the spectrum will be shown if P = 2N. SASE-V smooths I(k) over 5, 10, and 20 points using a rectangular window, and plots the results.

Theoretically, the spectrum for a homogeneous Poisson process would be flat.

REFERENCES

1. Cox, D. R. and P. A. W. Lewis, 1966, *The Statistical Analysis of Series of Events,* Methuen, London.

2. Durbin, J., 1961, "Some Methods of Constructing Exact Tests," *Biometrika*, Vol. 48, pp. 41-55.

3. Kendall, M. G. and A. Stuart, 1961, *The Advanced Theory of Statistics*, Vol. 2, pp. 450-461. Hafner Publishing Co., N. Y.

4. Lewis, P. A. W., 1961, "Distribution of the Anderson-Darling Statistic," *Annals of Mathematic Statistics,* Vol. 32, pp. 1118-1123.

5. Lewis, P. A. W., 1965, "Some Results on Tests for Poisson Processes," *Biometrika*, Vol. 52, pp. 67-77.

6. Lewis, P. A. W., A. M. Katcher, and A. H. Weis, 1970, "SASE-IV – An Improved Program for the Statistical Analysis of Series of Events," Program Order Number 360L - 13.0.001, IBM Corp., Program Information Department, 40 Saw Mill River Rd., Hawthorne, New York 10532.

APPENDIX B
CANDIDATE DENSITIES

The Gamma Density:

The gamma density is

$$f(x) = \left(\frac{r}{u}\right)^r \frac{1}{\Gamma(r)} x^{r-1} e^{-rx/u} h(x)$$

where h (x) is the unit step function and

$$\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt$$

is the gamma function. Notice that for $r = 1$, the gamma density reduces to the exponential density. The expected value for a gamma random variable X is $E\{X\} = u$, the variance is $\sigma^2 = u^2/r$, and the coefficient of variation is $C = \sigma/u = r^{-1/2}$.

The log likelihood function for a random sample is

$$\ln f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i)$$

$$= n\, r \ln \frac{r}{u} - n \ln \Gamma(r) + (r-1) \sum_{i=1}^{n} \ln x_i - \frac{r}{u} \sum_{i=1}^{n} x_i$$

The maximum likelihood estimates for r and u satisfy

$$\frac{\partial}{\partial r} \ln f(x_1, \ldots, x_n) = 0$$

and

$$\frac{\partial}{\partial u} \ln f(x_1, \ldots, x_n) = 0$$

From the second equation we find that the maximum likelihood estimate for u is

$$\hat{u} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Substituting $\hat{u}$ into the first equation, we find that the estimate for r must satisfy the transcendental equation

$$\ln \hat{r} - \psi(\hat{r}) = \ln \hat{u} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

where

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x)$$

is the digamma function.

The Rayleigh Density:

The Rayleigh Density is

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} h(x)$$

In this case $E\{X\} = \sigma\sqrt{\pi/2}$ and var $x = \sigma^2 (2 - \pi/2)$. The log likelihood function for a random sample is

$$\ln f(x_1, \ldots, x_n) = -n \ln \sigma^2 + \sum_{i=1}^{n} \ln x_i - \frac{2}{\sigma^2} \sum_{i=1}^{n} x_i^2$$

Differentiating with respect to σ^2 and setting the result to zero, the maximum likelihood estimate is found to be

$$\hat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^{n} x_i^2$$

It can be shown that this estimate is unbiased, i.e., $E\{\hat{\sigma}^2\} = \sigma^2$, and has the variance σ^4/n.

The Lognormal Density:

The lognormal density is

$$f(x) = \frac{1}{x\,\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(\ln x - m)^2}{2\sigma^2}\right] h(x)$$

The random variable $Y = \ln X$ is normally distributed with mean m and variance σ^2. The log likelihood function for a random sample is

$$\ln f(x_1, \ldots, x_n) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\ln x_i - \frac{2}{\sigma^2}\sum_{i=1}^{n}(\ln x_i - m)^2$$

By setting the partial derivatives with respect to m and σ equal to zero, the maximum likelihood estimates are found to be

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n}\ln x_i$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \hat{m})^2$$

APPENDIX C
Lightning Data and Computer Simulation

Table 1 presents the results of tests made on data from four storms in August and September, 1975 near Atlanta and in July 1976 during TRIP at the Kennedy Space Center. The column entitled "exponential" presents the results of a uniform conditional test on the time intervals, testing for Poisson statistics. The statistic, KS, is the result of the test without Durbin's transformation, and DN is the statistic for the test with Durbin's transformation. The histogram of some of this data was also tested against candidate density functions (see Appendix B). The results of these tests are given in the last three columns. The significance of the statistic, KS or DN, is given in Table 5.

Table 2 presents results of the simulation. Data were simulated assuming an exponentially distributed time between events (τ). In all examples shown in Table 2 the mean interval between events is one second. The duration (σ) of each event was modeled as a constant (c) plus an exponentially distributed random variable with rate parameter λ. The mean duration is $\langle\sigma\rangle = 1/\lambda + c$. In Data Set A the constant is zero, and in Data Set B the duration was a constant, c, only. The size of the constant is varied in the remaining examples. Each set of simulated data was tested against the three density functions shown: Gamma, Rayleigh and Lognormal. The significance of the KS statistic derived in this test is given in Table 5.

Table 3 presents additional results of the simulation. In Table 3, data are presented in which the time between events is varied. In these examples the time between events is exponentially distributed with durations which are modeled as a constant, c, plus an exponentially distributed random variable. The rate parameter, λ, of the random variable and the constant are fixed such that $\lambda c = 1$ and $c = 1/6$. This corresponds to a mean duration of 1/3 second. The first row in this data set pertains to events which are on the average 8 seconds apart (a low flashing rate) and the last row corresponds to events which are on the average 1 second apart (a high flashing rate). Notice that the lognormal density fits best in the case of the high flashing rate where overlaps are most important.

In Table 4, results are shown for an example where nonexponential statistics were used for the time between events. The data here were obtained assuming that the time between events had a density function of the form $\exp(-at^2)$ with a mean interval of 1 second. The duration was a constant plus an exponentially distributed random variable as above. Notice that the lognormal density is rejected for this case. Other nonexponential functions were also tried for the distribution of the time intervals with similar results.

Table 5 shows significance levels for the Kolmogorov-Smirnov statistic, KS.

Table 1
Lightning Data

Storm Date	Tape Number	Sample Mean	Sample Variance	Exponential KS	Exponential DN	Gamma KS	Rayleigh KS	Lognormal KS
8/5/75								
	1	3.50	3.00	0.63	1.40			
	2	2.90	2.30	1.05	2.64			
	3	1.80	1.60	0.70	3.32			
	4	1.40	1.40	0.89	2.20			
	5	6.10	6.90	2.90	2.00			
8/26/75								
	1	0.86	0.50	0.72	8.23	1.70	3.10	1.93
	2	1.05	0.64	0.81	9.12	6.50	3.50	1.54
	3	1.48	0.87	0.56	6.68	0.96	2.00	0.86
	4	1.74	1.02	1.19	7.92	1.02	2.30	1.47
9/12/75								
	1	2.40	1.60	0.68	4.50	1.18	2.60	0.67
	2	1.20	0.80	1.18	9.02	12.3	3.60	1.54
	3	1.70	1.10	0.79	4.82	0.94	2.70	1.23
	4	1.80	1.30	1.43	6.30	1.49	4.30	1.01
	5	2.40	1.70	0.84	2.89	0.85	2.40	1.30
7/13/76 (all flashes)								
	1	1.91	1.20	2.09	8.46	57.1	2.67	0.83
	2	1.94	1.30	1.44	7.30	1.51	3.64	0.71
	3	2.32	1.50	1.24	6.63	1.66	3.59	0.82
(cloud-to-ground flashes)								
	1	20.5	17.6	1.48	0.76			
	2	10.3	08.7	0.70	1.63			
	3	17.0	14.8	2.00	1.94			

KS = Kolmogorov-Smirnov statistic
DN = Durbin's transformation

Table 2
Computer Simulations for Exponential Time Interval

Mean Duration			Sample Mean	Sample Variance	Gamma KS	Rayleigh KS	Lognormal KS
<σ>	λ^{-1}	c					
Data Set: A (Exponential duration)							
0.50	0.50	0	1.67	1.54	0.63	6.98	2.65
0.33	0.33	0	1.37	1.10	0.65	7.67	2.43
0.25	0.25	0	1.29	1.12	0.85	8.80	2.31
0.20	0.20	0	1.23	1.06	1.24	9.83	2.10
Data Set: B (Constant duration)							
0.50	00	0.50	1.67	1.01	26.40	4.50	1.60
0.33	00	0.33	1.39	1.02	3.14	7.89	1.50
1.25	00	0.25	1.27	0.96	5.45	8.40	1.20
0.20	00	0.20	1.22	0.95	2.57	9.40	1.80
Data Set: C ($\lambda c = 1/2$)							
0.50	0.33	0.17	1.65	1.12	6.48	5.60	1.00
0.33	0.22	0.11	1.39	1.13	21.80	7.90	1.20
0.25	0.17	0.08	1.28	1.01	1.91	8.90	1.14
0.20	0.13	0.07	1.22	0.98	2.03	9.70	1.10
Data Set: D ($\lambda c = 1$)							
0.50	0.25	0.25	1.66	1.18	64.90	5.45	0.67
0.33	0.17	0.17	1.39	1.14	481.30	7.95	0.90
0.25	0.13	0.13	1.29	1.01	2.34	8.96	1.36
0.20	0.10	0.10	1.22	0.99	2.27	9.76	1.05
Data Set: E ($\lambda c = 2$)							
0.50	0.17	0.33	1.67	1.26	173.00	5.30	1.04
0.33	0.11	0.22	1.40	1.20	1916.00	7.90	1.14
0.25	0.08	0.17	1.29	1.03	2.50	9.20	1.46
0.20	0.07	0.13	1.21	0.99	25.00	9.70	1.22

Exponential interval for all cases
Mean interval = 1 sec.
Mean duration = $\lambda^{-1} + c = \langle\sigma\rangle$

Table 3
Computer Simulation for Variable Interval

Mean Interval	Mean Duration			Sample Mean	Sample Variance	Gamma KS	Rayleigh KS	Lognormal KS
	$<\sigma>$	λ^{-1}	c					
8	0.33	0.17	0.17	8.45	0.60	1.08	11.70	2.40
4	0.33	0.17	0.17	4.46	17.20	1.86	12.10	2.08
3	0.33	0.17	0.17	1.72	1.79	2.34	8.90	1.35
1	0.33	0.17	0.17	1.36	0.98	2.42	7.60	0.99

Exponentional time intervals
Variable mean interval
Duration fixed = 0.333 sec.
$\lambda c = 1$
Mean duration = $\lambda^{-1} + c = 1/3$

Table 4
Computer Simulation for Nonexponential Duration

Mean Duration			Sample Mean	Sample Variance	Gamma KS	Rayleigh KS	Lognormal KS
$<\sigma>$	λ^{-1}	c					
0.50	0.25	0.25	1.48	0.52	219.00	1.90	2.00
0.33	0.17	0.17	1.23	0.47	1.14	2.20	1.90
0.25	0.13	0.13	1.22	0.55	0.98	3.60	1.75
0.20	0.10	0.10	1.17	0.47	0.99	2.90	2.64

Time intervals density of the form $e^{-\alpha t^2}$
Mean durations = $\lambda^{-1} + c$
Mean interval = 1 sec

Table 5
Asymptotic Significance Levels for the
Kolmogrov-Smirnov Statistic

z	$P(D_N > z)$
.10	.10000000+01
.15	.10000000+01
.20	.10000000+01
.25	.99999999+00
.30	.99999070+00
.35	.99969714+00
.40	.99719233+00
.45	.98741063+00
.50	.96394525+00
.55	.92281682+00
.60	.86428282+00
.65	.79201308+00
.70	.71123525+00
.75	.62716711+00
.80	.54414248+00
.85	.46531929+00
.90	.39273078+00
.95	.32748555+00
1.00	.26999973+00
1.05	.22020562+00
1.10	.17771825+00
1.15	.14195991+00
1.20	.11224971+00
1.25	.87866448-01
1.30	.68092250-01
1.35	.52241911-01
1.40	.39681899-01
1.45	.29841489-01
1.50	.22217975-01
1.55	.16377412-01
1.60	.11952054-01
1.65	.86356858-02
1.70	.61774354-02
1.75	.43749856-02
1.80	.30676239-02
1.85	.21295343-02
1.90	.14636061-02
1.95	.99591177-03
2.00	.67092590-03
2.05	.44749213-03
2.10	.29549713-03
2.15	.19318713-03
2.20	.12504322-03
2.25	.80130748-04
2.30	.50838800-04
2.35	.31933641-04
2.40	.19859059-04
2.45	.12227169-04
2.50	.74533286-05
2.55	.44981264-05

1. Report No.
NASA TP-1665

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle

RATE STATISTICS FOR RADIO NOISE FROM LIGHTNING

5. Report Date
May 1980

6. Performing Organization Code
Code 946

7. Author(s)
D. M. Le Vine, R. Meneghini and S. A. Tretter

8. Performing Organization Report No.
G80F5117

9. Performing Organization Name and Address

NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

10. Work Unit No.

11. Contract or Grant No.

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, D.C. 20546

13. Type of Report and Period Covered

Technical Publication

14. Sponsoring Agency Code

15. Supplementary Notes

D. M. LeVine and R. Meneghini, Goddard Space Flight Center.
S. A. Tretter, Department of Electrical Engineering, University of Maryland, College Park, Maryland.

16. Abstract

Radio frequency noise from lightning was measured at several frequencies in the HF – VHF range during the Thunderstorm Research International Project (TRIP) at the Kennedy Space Center, Florida. The data were examined to determine flashing rate statistics during periods of strong activity from nearby storms. It has been found that the time between flashes is modeled reasonably well by a random variable with a lognormal distribution.

Initially, the hypothesis that the occurrence of lightning flashes is a Poisson point process was tested using a uniform conditional test with Durbin's transformation and the Kolmogorov-Smirnov statistic. However, the Poisson hypothesis failed for the measured data. This resulted mainly from a lack of small time intervals, which we believe is a consequence of the finite duration of lightning flashes. This hypothesis has been tested by simulation, using a compound process in which the intervals between lightning flashes is assumed to be exponentially distributed and each flash is assumed to have a duration given by another independent random variable. Assuming flash rates and durations consistent with data, and that the flash duration has a fixed minimum of the order of the measuring system's resolution, the lognormal distribution consistently fit the simulated data.

17. Key Words (Selected by Author(s))
Radio Noise, Lightning, Flashing Rate Statistics, TRIP

18. Distribution Statement

STAR Category 47
Unclassified – Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	33	$4.50

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1980